

DE-GOLDENIZING KAZAKHSTAN'S GOLD AND FX ASSETS

BY
SANZHAR BERIK

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Finance
in the Graduate School of Business
Nazarbayev University, 2025

Astana, Kazakhstan

Advisor: Dr. Francesco Rocciolo

Abstract

This thesis is dedicated to the analysis of the market risk of the gold and foreign exchange assets of Kazakhstan and the evaluation of the extent to which their current structure is justified, which is also characterized by an abnormally high share of gold. The study aims at determining the actual magnitude of possible losses and how various risk models can capture the behavior of portfolios when the market is relaxed and when it is stressed.

The analysis is made on the basis of replication of reserves with references to the market indices and to the application of a contemporary set of techniques: historical modeling, dynamic GARCH models, and a modified Expected Shortfall indicator that satisfies the Basel III requirements. The quality of the risk forecasts was also tested using backtesting and a regime switching model, which determines the shifting stages of the financial cycle and its influence on the structure of the optimal portfolio. The obtained results prove that the distribution of reserve yields are those with pronouncedly heavy-tailed and that the traditional approaches tend to smooth out the real risks. The dynamic models will give a more realistic picture of how the portfolio will respond in times of market turmoil, showing that the tail losses will rise dramatically. The regime analysis reveals that the gold does enhance the stability of the portfolio when stressful times set in, but it is a high-cost and loss-of-profit source when the growth is in a stable state.

This work identifies the relevance of an adaptive and market-sensitive solution and illustrates the weakness of a fixed reserve framework in a highly dynamic global financial system.

Keywords: Expected Shortfall · GARCH · Historical Simulation · Basel III · Markov Regime Switching · Tail Risk · GXFA · Gold Allocation

1. Introduction

The structure of international reserves is significant in the sustainability of the country's macroeconomic sphere and preservation of confidence in the national currency, and minimizing exposure to exogenous shocks. In the case of Kazakhstan, the management of the reserves is of special significance because the economy of the country is still subject to the influence of the commodity prices on the international markets, the processes on the international financial markets, and the flow of capital. Given the high volatility of the oil market, cyclical geopolitical crises, and modification of the monetary policy of the developed countries, the structuring of gold and foreign exchange assets (GFXA) requires an accurate and scientifically determined approach. Based on the official details of the National Bank of Kazakhstan, the proportion of gold in the reserve portfolio of the country has always been well above 50 percent over the past years, which is much higher than in other similar economies where gold is usually limited to 10-20 percent. Such a level of concentration is not normal because Kazakhstan is among the few nations that have a heavy reliance on one type of asset in its reserve structure. The figure below shows the current composition of the reserve portfolio with a strong focus on gold as compared to investment, liquid and strategy, and other alternative assets.

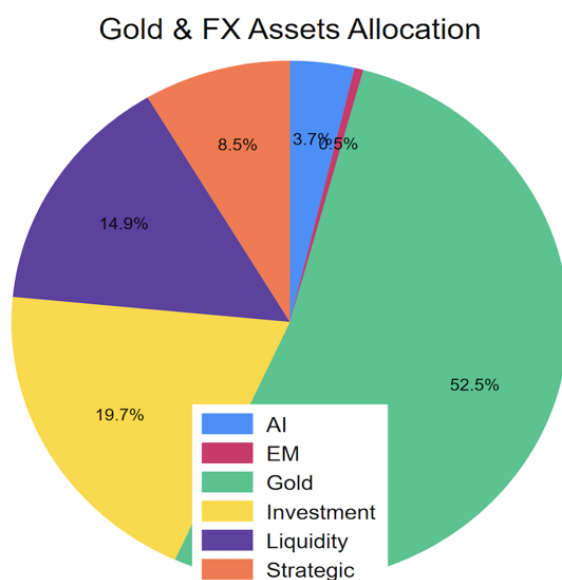


Figure 1. Total portfolio allocation of the NBK's GFXA (end-2024)

Having a high percentage of gold in the portfolio can at the same time be a source of strength and a source of risk. Firstly, the historical understanding of gold is that it is a safe-haven asset, which

makes reserve stability stronger during times of uncertainty in the world. Conversely, recent research indicates that safe-haven properties of gold are not absolute (Baur & Lucey, 2010): they are subject to market regimes, shock nature, liquidity in the global markets, as well as interest rate dynamics. In calm market conditions, gold can be highly volatile and gain tail risk, whereas other instruments like the bonds of the U.S. Treasury, or the bonds of international financial organizations will have a more predictable profile of returns.

This makes determining the most efficient reserve structure especially significant to an economy whose financial markets are underdeveloped, and the rate of commodity reliance is high, as in the case of Kazakhstan. As markets shift between stable and stressed regimes, the traditional reserve management practices might underestimate risk and lack the ability to consider the variation in the behavior of the assets in different market conditions.

The academic literature has gained a lot of work on the study of gold as a safe-haven asset and measures the market risk with the assistance of new statistical tools. Research indicates that risk measures calculated through Value-at-Risk (VAR) are not effective enough to measure tail risk, and their efficiency decreases when financial markets undergo regime changes. Basel III/IV has, in turn, responded by making the Expected Shortfall (ES) its main market risk measure.

When applied to the situation in Kazakhstan, the analysis of the dependence of the market risk of reserves on the portfolio composition is of particular interest. Since the proportion of gold is high, it is important to know when exactly it is used as a safe-haven asset, its role in the extreme losses of the portfolio, and whether there is an economically justified maximum percentage of gold in the reserves. These questions characterize the main issue discussed in this paper.

Problems and research questions

With the emerging academic literature on the properties of gold and the approaches to risk assessment, the issue of whether the gold concentration in the reserves of Kazakhstan is economically justified has not been adequately addressed. One of the challenges is to understand how the gold will behave in various market regimes, as its effectiveness as a safe-haven asset will eventually be determined by the underlying regime structure.

Based on this, the main research questions are formulated as follows:

1. Is gold a reliable safe-haven asset for Kazakhstan, or does its high share increase the vulnerability of reserves under different market regimes?

2. What is the opportunity cost of holding more than 50% of reserves in gold compared with holding currency-denominated assets such as the U.S. Treasury securities, bonds of supranational institutions, and other instruments?

These questions are complemented by several more specific ones:

- To what extent does gold increase the Expected Shortfall of the reserve portfolio?
- How does the contribution of different GFXA components to tail risk change?
- What is the optimal share of gold given the potential market transition to a stressed regime?
- How do regime-dependent risk models compare with classical methods?

Motivation

The motivation of this study arises from the fact that the traditional approach to managing central banks' reserves becomes less applicable in the modern, highly volatile financial environment. Global markets for the past ten years have been characterized by non-constant short stable periods, which are then followed by extended periods of uncertainty, which are driven by changes in monetary policies of central banks, geopolitical tensions among nations, high fluctuations in commodity prices, and finally inflation (IMF, 2020). For emerging markets, such as Kazakhstan, which are dependent on the export of materials integrated into global financial flows, the swings in this magnitude lead to a high risk to the stability of reserves and require a qualitatively different system.

The distinctive feature of Kazakhstan is the established structure of GFXA, in which gold occupies a dominant position. From a practical viewpoint, this approach seems rational and at the same time vulnerable: gold is traditionally considered a long-term asset, but modern empirical literature shows that its behavior significantly depends on the state of the market. It might be resilient during crisis periods, but it loses its features during a calm macroeconomic regime. Hence, high concentration in one asset leads to asymmetrical reaction of the portfolio on market condition changes, and this asymmetry, in particular, has long lacked a formal and systematic evaluation.

The second aspect that justifies the motivation of this research is the necessity of revision the risk assessment methodology. Over the last years, international regulators have emphasized the importance of shifting from risk measures that rely purely on distributional assumptions of returns to more flexible and informative risk metrics. In particular, ES, which is recommended by the Basel Committee, is able to reflect the tail behavior distribution. In addition, the very nature of the

reserve portfolio's risk, especially with a high share of gold, cannot be properly described without incorporating market-regime shifts.

This leads to a key methodological challenge: standard models based on historical return distributions cannot capture the probabilistic structure of transitions between phases of calm and stress periods. This makes the application of regime-switching models, such as the Markov-switching model, not only reasonable but effectively essential. Only those kinds of models are able to assess to what extent the risk of reserves might change if the economy were in a turbulent market situation, and these models can finally propose what the optimal weight of gold is in possible regimes.

Finally, the practical main motivation of the study arises from the lack of a comprehensive assessment that would combine information about market regimes, volatility levels, the structure of the portfolio, and the impact of individual components on extreme tail losses. Before this research, there were no similar practical studies that would consider the relationship between the high share of gold in Kazakhstan's GXFA and the regime dynamics of risk. It is unclear to what extent current assets' composition is resilient to a long-term perspective, what alternative structures would ensure a more balanced risk profile, and what role gold plays in constructing a tail risk portfolio.

To conclude, the motivation of this research is the necessity to provide a quantitative answer to strategically important questions: how justified is the current concentration of gold in GXFA, how does it affect the stability of the portfolio across different regimes, what finally what structural adjustments might decrease the likelihood of extreme losses. The proposed study aims at bridging the gap by providing a thorough regime-specific analysis of risk modified to the characteristics of the Kazakhstan financial system.

The purpose and objectives

This research aims to give a detailed evaluation in regard to the influence of the structure of the reserve portfolio of the National Bank of Kazakhstan on the market risk under the modern analytical techniques, such as Expected Shortfall, dynamic volatility models, and regime switching models. In order to reach this objective, the following are addressed:

1. Constructing a replication of the reserve portfolio of the National Bank of Kazakhstan, but using market proxy indices to represent all six elements.

2. Calculating Expected Shortfall and Value-at-Risk using the Historical Simulation and GARCH models.
3. Backtesting Value-at-Risk and Expected Shortfall in full, including the measurement of violations, actual tail losses, and model risk.
4. Using a two-regime Markov Switching model to determine calm and stressful conditions of the financial market.
5. Computing regime-specific estimates and calculating the optimum amount of the gold investment with reference to the prevailing and expected market regime, determining the opportunity cost of a high gold share by comparing the profiles of various GFXA elements.

Scientific and practical significance

This study contributes to the literature on reserve management and tail risk modeling as it combines the various methods that are seldom used together. It focuses on Expected Shortfall as the key measure of risk, as consistent with Basel III, it uses dynamic volatility and regime models which capture structural changes in financial markets, and it uses backtesting-based ES adjustments which give reasonable estimates of tail losses, and focuses on optimization of reserve structures based on risk-sensitive measures, not just returns or volatility.

The practical importance of the given piece of work is the following: the obtained results may be used to conduct the quantitative evaluation of the risks that may be related to the high volume of gold share, to establish the framework of the adaptive reserve management strategy development, to illustrate the necessity of regime-specific approach to the GFXA structure, and to estimate the reasonable range of the gold share based on the market state.

2. Literature Review

According to the recent research on the composition of international reserves and the management of their risks, a thorough analysis of assets that represent a substantial portion of central bank portfolios should be done. In Kazakhstan, the asset is gold, and the concentration of this asset is considerably more than in most other countries with such economic characteristics. The high level of gold concentration supports the need to measure the related risks properly, and also identify the dependence of the risk of this precious metal on the financial markets of the world.

The academic literature, which has been accumulated over the past few years, has taken two major directions that are directly applicable to this study. The first direction is associated with the study of the nature of gold as a safe-haven asset: it has been discovered that the functions of gold are heterogeneous and depend on the situation in the economy business cycle, and global volatility. The second direction looks at the measurement of the risk of assets with heavy distribution tails. The shift from Value at Risk to Expected Shortfall is particularly important because ES reflects the extent of possible losses in times of stress. Nevertheless, such a change is also accompanied by the fact that risk assessment becomes much more sensitive to the type of statistical model used, which leads to the appearance of the so-called model risk.

So, the theoretical framework of the research is based on scientific research over the past 10 years, which is a combination of studying the behavior of gold, market regime, and measuring the risks in terms of Expected Shortfall.

Historically, gold has been viewed as a relatively stable component of reserve reserves, which would act as a natural shock absorber against inflation and financial crisis. The development of the empirical techniques and long-term data has, however, allowed the researchers to investigate the behavior of gold further and to disprove the idea of its universal defense role.

Beckmann, Berger, and Czudaj (2019) have conducted one of the most important studies on gold dynamics. The behavior of gold is not described as either monotonically protective or constantly hedging, as demonstrated by the authors. Rather, gold shows strong regime dependence: during the times of global crises and of extreme market volatility, it, in fact, performs as a safe-haven asset, but when times are calm, and the markets are transitional, this behavior is likely to lose its defensive characteristics and even become speculative. A notable conclusion of their research is the fact that gold does not respond to the variation in market uncertainty in an equal way, and the

pattern of its relations with the main macroeconomic indicators - real interest rates and perceived market risk - differs greatly according to the existing market regime (Baur & McDermott, 2010). This method is a challenge to the traditional linear view of gold as a universal hedge. Rather, the authors propose to consider gold as an asset whose performance was determined by the global economic environment. In particular, the global business cycle stage, the cost of risk relative, and the macroeconomic expectations dynamics. This causes gold not just to be a good diversifier, but also provides a serious regime risk, particularly when a large proportion of government reserves is held in gold.

Such findings are especially significant from the perspective of Kazakhstan. The skewness of gold in the reserve structure means that the strength of the total portfolio is highly determined by the quality of the delimitation of the economic regimes and the degree of accuracy of the risks associated with the economic regimes. The risk of an incorrect identification of the regime may result in either an incorrect evaluation of the reserve resilience or an underestimation of vulnerability, as well as inappropriate tactical decisions. In that regard, the article by Beckmann et al. offers a conceptual justification to integrate models that reflect the structural changes and regime shifts into the current study, as well as to note that analytical care is to be taken when interpreting the price of gold.

The traditional market risk measurement used to be based on the Value at Risk metric. Nevertheless, the first of the limitations that have been recognized in various critical works regarding VaR is that it fails to reflect the importance of losses that appear at the level of the selected confidence, and, as a result, it does not reflect the risks that occur under extreme circumstances. After the global financial crisis, regulators slowly came to the conclusion that VaR was not giving the financial system enough resilience, and this, eventually, resulted in the replacement of VaR by Expected Shortfall in the Basel standards.

Unlike VaR, Expected Shortfall calculates the average worst-case losses and is more tail-sensitive. It makes ES a more relevant instrument in the analysis of stress events and the extreme financial instability paradigm.

Lazar and Zhang (2019) explicitly discuss model risk problems, which is a significant research topic in the field of Expected Shortfall estimation. The authors examine the sensitivity of ES accuracy to the distribution, model structure, and volatility assumptions. Using broad simulations,

they are able to conclude that the sensitivity of ES to model specification is far greater than many practitioners initially anticipated.

Particular attention is given to the fact that the errors, arising from an incorrect choice of return distribution, turn out to be much more significant than the errors associated with the estimation of the parameters within the chosen model. This means that Expected Shortfall can be seriously biased if a practitioner assumes a normal distribution, whereas the actual data are characterized by heavier tails. The authors emphasize that a normal model systematically underestimates the probability of extreme losses and, therefore, produces an overly optimistic ES estimate.

Also, the paper illustrates that time-varying volatility models like GARCH have the ability to reduce the extent of risk underestimation, although it does not ensure the absence of the latter in case the original assumptions of tail distributions are still misleading. Lazar and Zhang also propose approaches for backtesting of ES that allow for a more accurate evaluation of model robustness under conditions of limited observations, which is particularly important when analyzing rare stress events.

Those results are very important in the asset management context. A high level of gold concentration means one is highly exposed to tail risk, and this simply implies that model uncertainty must be carefully managed when using ES. ES can only be used properly in the event that the model used captures the volatility, as well as the structure of the tail distribution of gold and other assets.

The literature shows that gold cannot be regarded as a safe asset that, with regard to any changes in market conditions, is resilient. It is organized according to structural changes, regime-specific volatility, and responsiveness to the global economic situation, which have been reported by Beckmann et al (2019). At the same time, the risk measurement of gold requires the application of ES as the most sensitive metric, though these models possess a high level of model risk, which is shown by the work of Lazar and Zhang (2019).

The combination of both directions leads to the need to construct regime-oriented systems of risk assessment, which makes it possible to adapt to any changes in the features of assets and correct the reserve structure, considering actual macroeconomic conditions. The methodological foundation of this thesis is grounded exactly at the intersection of gold's regime-dependent dynamics and the model risk of ES.

3. Data

In the empirical part of my thesis, two datasets are used: (1) daily returns of market indices that are used for constructing synthetic returns of the National Bank of Kazakhstan's (hereafter, NBK) Gold and FX Assets (hereafter, GFXA), and (2) macro-economic variables for computing the Markov-switching model. Daily returns were collected from the Bloomberg database, FRED, and further combined into several Excel source files. The time span of the analysis covers from 01/04/2021 to 08/29/2025. This panel includes market prices of gold, stock, and bond indices of developing and developed nations, money market indicators, as well as the fiat currencies' exchange rates.

The structure of the data allows for the construction of the market dynamics of each component of the GFXA. In particular, the gold portfolio is described by three proxy instruments: physical gold, ETF GLD, and a gold derivative - together, they reflect the structure of the gold assets.

Liquidity portfolio is represented by indicators of the money market and short-term government bonds: the Cash Index, one-month US T-Bills, US agency bonds, and, finally, US short-term Treasury Bonds.

Strategic portfolio includes global investment-grade bonds and its riskier corporate segment, along with sovereign and corporate debt instruments of developing countries.

An investment portfolio reflects the debt and currency structure of reserves. This portfolio is represented by a series of government bonds of developed countries: the US, the UK, Australia, Canada, and the eurozone.

Alternative Investment portfolio (AI) covers the Global Equity Index, the Public Real Estate Index, the Liquid Alternative Index, and the IG bond Index.

Emerging Market portfolio is formed by solely Chinese Yuan, which includes Chinese government bonds, local instruments in CNY, and the currency component measured through the daily change in the CNY/USD exchange rate.

After collecting all components, the total returns of the aggregated GFX are calculated. Finally, those returns are used as an input for VaR and ES calculations.

In addition, macro-economic variables are used for identifying market regimes, such as economic slowdown and expansion, which are then used for the Markov-switching model.

The above-mentioned economic variables could be divided into two sets. First, daily returns of external macroeconomic variables, such as oil prices (futures for Brent Oil), yield on 10-year US

T-Bill, and the short-term US interest rate. Those factors reflect the global risk, the current oil market condition, and the global direction of interest rates, which influence the market volatility. Moreover, internal indicators are used (e.g., index “kase” – the indicator of internal market index, and GDP growth of Kazakhstan, which reflects the state of the real economy), which allow for the extension of the model’s standard risk analysis to a regime-based framework, accounting for both external and internal factors of Kazakhstan’s economic environment. Full data is shown in Table 1.

Portfolios	Sector Allocation:	
	NBK rules	Constructed proxies
Gold	Unallocated gold – 90-100% Gold ETFs – 0-10% Derivatives 0-10%	Gold – 90% Gold ETF – 5% Gold derivatives – 5%
Strategic	Global Corporate Bonds (ICE BofA Global Corporate & High Yield Index) – 50% Emerging Market Debt (ICE BofA EM indices) – 50%	IG corp – 45% HY corp – 5% EM sov – 26% EM corp – 24%
Investment	Sovereign debt securities/deposits: Currency basket: USD – 62%; EUR – 12.5%; GBP – 12.5%; AUD – 8%; CAD – 5%	USD – 62% EUR – 12.5% GBP – 12.5% AUD – 8% CAD – 5%
Liquidity	Cash (FX) and sov. debt – 70-100% Agency debt – 0-30% Certificates of Deposit/CPs - 0-30%	Sov. Debt – 70% Cash – 17% Agency debt – 8% Deposits – 5%
Alternative Investments	Global Equity index – 80% Global Bond index – 20% Sectors: PE, HF, P, and Bonds	Private Equity – 37% Property – 28% Hedge Fund – 30% IG corp – 5%
Emerging Market	Sovereign debt securities, Assets denominated in CNY – 100%	CNY sov. bonds – 70% CNY policy bank – 20% CNY cash – 10%

Table 1: Portfolio structure of NBK’s GFXA

4. Methodology

The methodological part of the research is directed to construct a proper and reproducible framework for assessing the market risk of NBK's GFXA, taking into their actual structure, the dynamics of global financial markets, and the requirements of international risk standards (in particular, Basel III). The comprehensive methodological scheme was constructed for achieving this aim, which includes:

1. The construction of a real portfolio based on market proxy indices.
2. Ensuring the construction of homogeneous and high-quality time series of returns.
3. Assessing the risk based on a series of models (Historical Simulation, GARCH, ES).
4. Scaling the risk measure to longer horizons.
5. Backtesting
6. Modeling regime switching to account for market condition instability.

This comprehensive approach ensures analytical completeness and allows us to obtain a consistent picture of the risk behavior during both normal periods and stressful market conditions.

The construction of a synthetic portfolio of GFXA

The NBK does not publicly publish its daily market estimates of the profitability of each segment of GFXA. Therefore, in order to model risk, it is necessary to reproduce the dynamics of assets through a set of publicly observable instruments that reflect the portfolio structure. This method is widely used in practice in the literature on reserve management (IMF, 2003), where the benchmark replication method is used – the construction of a synthesis series of returns through market indices, which represent the risk profile of the underlying portfolio. In this paper, for each sector of GFXA, a separate proxy basket is selected, and this is aggregated into a single total portfolio, which allows us to reproduce the dynamics of assets. The indices have high liquidity with a long history of observations and correspond to assets of NBK's GFXA. Thele 1 (Portfolio structure of NBK's GFXA) reflects the relationship between the official rules and constructed market analogues.

Next, logarithmic returns are formed from the closed prices according to the formula:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right),$$

which were further used in the construction of portfolio returns and for VaR/ES models.

Every sector is a linear combination of the returns of the component proxies

For example, for the gold sector:

$$r_{gold,t} = 0.90r_{cash,t} + 0.05r_{ETF,t} + 0.05r_{deriv,t}.$$

Finally, a single row of profitability of GFXA is created:

$$r_{total,t} = \sum_{i \in \{gold, liq, str, inv, ai, em\}} w_i r_{i,t}.$$

This series reflects the dynamics of the whole asset portfolio and is used at subsequent stages of risk analysis.

Value-at-Risk (VaR)

The assessment of potential losses, which might occur under unpleasant changes of market conditions, is the central element of the risk analysis of the central bank's reserves. Among many approaches of risk assessment models, the one that stands out the most is Value-at-Risk (VaR), which takes the primary position due to its plain approach in interpretation, along with standardization in international banking practice and the possibility of application to a wide range of instruments. VaR shows the maximum potential loss of a portfolio at a given confidence level and over a fixed time horizon.

In this study, VaR is evaluated using two complementary methods:

1. Using the historical simulation method (HS-VaR), which is purely empirical and does not provide any distributional assumptions.
2. Using a parametric approach based on the conditional volatility model GARCH (1,1), which takes into account the clustering of volatility characteristic of financial time series.

Both methods reflect different aspects of the portfolio's risk profile, and their combined use ensures greater stability and completeness in assessing potential losses.

Financial time series possess two fundamental characteristics that make relying on a single VaR approach inadequate. Firstly, empirical skewness and heavy-tailed distributions – returns on financial assets show significant deviations from the normal distributions. Under those conditions, methods based on parametric distributions may underestimate the profitability of extreme events. Secondly, volatility clustering – the periods of high volatility, as a rule, are followed by similar periods of high volatility, and vice versa. This makes the assumption of constant variance

unrealistic, and parametric models of conditional volatility (in particular, GARCH) more appropriate.

That's why in international practice, central banks generally use at least two methods of VaR to avoid bias in risk assessment. For example, BIS, ECB, and Bank of England state that the combination of HS and GARCH provides a more resilient risk profile (Bali, Mo, & Tang, 2008). Hence, this thesis uses both methods: HS-VaR reflects the actual empirical history of the portfolio, and GARCH-VaR takes into account the volatility and enables adaptation to rapidly changing market conditions. Their combination allows me to incorporate extreme historical events and capture current market conditions.

HS-VaR is one of the widely used non-parametric approaches, which is used by private financial institutions and regulators. Its key advantage is that it does not impose any distributional form, relying entirely on empirical observations. First of all, the window of historical data is determined by the length of N observations. In banking practice, 250-1000 business days are often used. Then, the actual returns for the period are collected inside the window $[t-N+1, t]$. All returns are sorted in ascending order: $r(1) \leq r(2) \leq \dots \leq r(N)$. Finally, the VaR value is defined as the negative value of the corresponding quantile:

$$VaR_{\alpha}^{HS}(t) = -q_{\alpha}(r_{t-N+1:t}).$$

This work accounts for 3 confidence intervals: 95% (standardized risk-management), 99% (a commonly used interval in banking supervision), and 97.5% (confidence interval, recommended by Basel III for Expected Shortfall). Whilst HS-VaR possesses clear advantages, such as high empirical realism since it is grounded in observed data, minimal model risk, and strong transparency, it has some disadvantages, e.g. the lack of responsiveness to volatility shifts (when markets move fast, HS adjusts with a significant lag) and dependence of the window size (if the span is too short, unstable estimates might occur, while a very long window smooths out extreme periods). Though HS remains a fundamental instrument for risk analysis.

The GARCH-VaR (Generalized Autoregressive Conditional Heteroskedasticity) model is the most popular method that has the feature of volatility clustering, which shows that periods of high fluctuations are followed by similar periods, and calm periods are followed by calm periods (McNeil & Frey, 2000).

The standard GARCH model (1,1) is set by the following system:

$$r_t = \mu + \varepsilon_t, \varepsilon_t = \sigma_t \zeta_t,$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where the model parameters ω, α, β are estimated using the maximum likelihood method.

After estimating the parameters, VaR is calculated as follows:

$$VaR_{\alpha}^{GARCH}(t) = -(\mu_t + \zeta_{\alpha} \sigma_{\alpha}),$$

Where ζ_{α} is the quantile of the normal or t-distribution.

The main advantages of GARCH-VaR are sensitivity to current market conditions – the model reacts to rapid changes in volatility, and the ability to forecast future volatility. However, there are some drawbacks, such as the sensitivity to misspecification of the error term and the risk of underestimating losses in the presence of extreme volatility shocks. In addition, there is a risk of failure to capture the nonlinear dynamics characteristics of regime-switching market environments. Nevertheless, GARCH remains the gold standard among models of volatility clustering (Bollerslev 1986).

Expected Shortfall (ES)

VaR answers the following questions: “What is the portfolio’s maximum loss given confidence level α ?” But it does not provide information about the magnitude of tail losses, which represents the fundamental drawback of the VaR measure. This issue was solved by the introduction of the Expected Shortfall model that became mandatory under Basel III (Lazar & Zhang, 2019).

ES has several advantages, such as accounting for heavy tails, which enables the possibility to capture the sensitivity of extreme events and captures the average loss given that outcomes fall below the VaR threshold, unlike VaR, which captures only a quantile. Additionally, ES is a coherent risk measure and far better reflects the pure nature of systemic risk, which is characterized by quite rare but extremely severe losses.

The ES model is used for assessing capital in international banks, and its application for risk assessment of banks’ reserves is methodologically right and modern.

$$ES_{\alpha} = -E[r_t | r_t \leq q_{\alpha}],$$

The following options of ES are computed in the analysis:

1. HS-ES (1-day) – historical ES;
2. GARCH-ES (1-day) – a forward-looking parametric ES;
3. Backtest-corrected ES – an enhanced version of ES based on backtesting errors;
4. Basel III 10-day ES – a scaled ES value which is commonly used in banking practice.

Basel Scaling (10-day horizon)

According to international standards, a 10-day horizon is required for assessing capital risk, which is caused by the need to account for the position liquidation period. Even during highly liquid portfolios (including reserves), the liquidation of large positions may require several days in a relatively stressed market environment. For scaling daily estimates, the following rule is used:

$$ES_{10} = ES_1\sqrt{10}, \quad VaR_{10} = VaR_1\sqrt{10}.$$

In the analysis, this rule is applied sequentially to three sets of assessment: (1) to HS - assessment of VaR and ES, (2) to GARCH-assessment, where volatility is forecasted based on model GARCH (1,1), and (3) to “backtest-corrected” ES, that is, adjusted ES estimates that take into account the backtesting results.

Backtesting of VaR and ES models

The backtesting procedure occupies a central role in the quality assessment and practical application of the market risk assessment models. Unlike theoretical features of VaR and ES, backtesting allows us to assert to what extent the model risk forecasts correspond to the actual portfolio losses observed in real historical data.

In this study, backtesting is carried out purely based on the time series of the daily returns of the GFXA’s portfolio sleeves. Methodological structure of backtesting system includes three interconnected elements: (1) checking the frequency of VaR breaches - exceedance rate, (2) assessing the magnitude level of actual tail risk – empirical ES and comparing it with initial model-based ES, and finally (3) adjusting ES based on observed breaches and computing model risk (Acerbi & Székely, 2014).

Unlike the classic minimal set of tests (Kupiec), an extended backtest approach is computed in the analysis, accounting for the behavior of VaR and the quality of the ES forecast. This approach correlates with the modern requirements of Basel, where the main confidence interval is 97.5%

The first step in backtesting is the exceedance rate. So, it is important to find out to what extent actual daily changes in the portfolio exceed the forecasted VaR 97.5%. That is, a binary indicator is formed for each trading day:

$$exceed_t = \begin{cases} 1, & \text{if } r_t < -VaR_t(97.5\%), \\ 0, & \text{otherwise.} \end{cases}$$

The mean of the indicator provides us with the empirical violation frequency. Ideally, at a 97.5% confidence interval, the expected breach frequency is 2.5%. Comparing the actual frequency with

the expected enables us to rate the pure correctness of the coverage. This backtest is equivalent to the Kupiec POF (Proportion-of-Failures) test, but in this analysis, it is applied in an expanded form.

Key elements of ES backtesting are the comparison of the model ES (predicted by the GARCH) and the actual average loss on the days when the VaR was violated. Empirically, the ES is described as:

$$ES_{emp} = -\frac{1}{N_{exc}} \sum_{t:r_t < -VaR_t} r_t,$$

This approach corresponds with the tail-expectation equation, and it is shown as a proper realization of backtesting, which is laid down under Basel standards.

The strongest side of the backtesting is the correction factor of ES. When the model underestimates the tail risk, the actual ES will be higher than the predicted one. This issue is checked through the coefficient:

$$corr_{factor} = \frac{ES_{emp}}{ES_{model}}$$

So, then the model underestimates the tail risk - $corr_factor > 1$ – ES rises. When overestimates $corr_factor < 1$ – ES is reduced. This step is a best practice and is found, for example, in BIS papers on the evaluation of ES models.

After correcting ES, it is scaled to the Basel III 10-day time-horizon. Then, we have the following formula:

$$ES_{corr}^{10d} = ES_{corr}^{1d} * \sqrt{10}.$$

The above-mentioned result is the primary model of market risk of the portfolio; the 10-day horizon is the foundation Basel III market-risk capital charge.

The study's code on Stata includes an important innovation: HS-ES is computed using a 252-day rolling window manually (without a built-in actual rolling operation) and is fully time-varying. So, methodology consists of the following steps: (1) a 2.5% percentile is calculated on each day (VaR 97.5%), (2) according to all observations, the average value is below this threshold, and (3) the average is taken with a negative sign, which is interpreted as the loss value. This ES is then scaled to a 10-day Basel time horizon. The primary aim is to ensure the comparison of HS-ES with GARCH-ES, which is used in model specification risk assessment.

Lastly, this analysis includes the rare block for academic papers: the model risk assessment, which consists of two components such as estimation risk and specification risk.

$$Estimation\ Risk = 100 * \frac{\sigma(ES^{Garch})}{\mu(ES^{Garch})}, \quad SpecRisk = \frac{\overline{ES^{HS}}}{\overline{ES^{GARCH}}}$$

When ES is rather volatile between sample periods, the model will be more sensitive to the sample, which results in high estimation risk. When it comes to specification risk, if spec_ratio > 1, then HS-ES is systematically higher, which makes ES reduce the risk.

Regime-switching models

One of the fundamental features of financial markets is their time-varying, non-homogeneous nature. Periods of low volatility, stable macroeconomic indicators, and pleasant external conditions often alternate with episodes of sharp price fluctuations, widening credit spreads, increased uncertainty, and commodity market declines. The traditional models with constant parameters (e.g., classic GARCH or HS with fixed time-horizons) assume that the statistical properties of returns remain the same during the entire sample. This might eventually lead to the following systematic issues: underestimation of risk during stressful periods, incorrect prediction of the models, overestimating the risk during quiet periods, false conclusions about the stability of the reserve portfolio, and ultimately, failure to take into account the influence of macro-economic factors.

Given the above-mentioned factors, one of the popular models is used in this study – Markov regime-switching models (hereafter, MSM). These models help to identify current market conditions of Kazakhstan based on daily observed data, to classify each date into a specific regime based on macroeconomic indicators, with GDP as the primary driver, to optimize the portfolio asset composition while taking into account regime-dependent asset dynamics, and finally combine financial and macro information within a unified decision-making system.

Unlike purely statistical hidden Markov models (MSM), this study uses a hybrid approach: it combines quantitative threshold rules, macroeconomic validation, and ultimately portfolio optimization. This design makes the model more transparent, interpretable, and resistant to swings in small samples.

The composition of indicators is used for constructing a regime model that reflects the internal environment of Kazakhstan and external financial conditions. The primary indicator is the KASE index, which reflects the market condition of Kazakhstan and is sensitive to internal investors' expectations. Besides, quarterly GDP data is added to the model, which is needed for macroeconomic verification of the correctness of the regime classification.

When it comes to external variables that traditionally impact the financial conditions of Kazakhstan, they are the Brent oil price since Kazakhstan is oil oil-exporting nation, and credit-default swaps of Kazakhstan, which reflect the premium for sovereign risk. CDS Kazakhstan is considered a proxy for volatility risk due to the fact that when sovereign risk rises, this index increases correspondingly. Additionally, the yield of 10-year US TIPS is included in the model to reflect global financial conditions along with US CDS.

For regime switching optimization, daily returns from six portfolio sleeves are used, which are then standardized to a daily frequency and missing observations imputed using the last available value. This approach corresponds to the principle of information accessibility: at any given time, the portfolio would be managed by exactly the set of data that was empirically published.

The composite risk index is calculated for each daym which increases if one of the following conditions is observed: (1) market index of Kazakhstan is in the lower quantile of the historical distribution, (2) brent oil is in the lower quantile, (3) CDS Kazakhstan exceeds historically high levels, and (4) real interest rate of TIPS have relatively higher values. Each signal reflects an individual source of total risk. Their combination forms the aggregated score, which shows to what extent the financial environment is unpleasant.

Once the parameters are set, the total risk index is converted into three regimes:

1. Expansion regime – low level of stress, relatively high level of Brent oil prices, and stable macro environment.
2. Slowdown regime – the beginning of deterioration of conditions, local shocks, and the rise of volatility.
3. Recession regime – simultaneous deterioration of several key indicators, rising sovereign risk, and falling market.

This approach helps to capture the economy's conditions in real time, hinging on solely available daily data. From a formal point of view, the correctness of the regime classification supposes the fulfillment of the following set of conditions in terms of conditional mathematical expectations:

$$E(GDP | Regime = Expansion) > E(GDP | Regime = Slowdown) > E(GDP | Regime = Recession).$$

This mathematical expectation means that in the expansion regime, expected GDP growth should be the highest, in the slowdown regime, moderate, and finally in the recession regime, minimal or close to zero. This formula reflects the basic principle of macroeconomic interpretation of regimes:

each regime captures a qualitatively different state of economic activity, and especially the difference in the conditional expectations of fundamental variables constitutes for validating the classification. Regimes that are constructed based on daily market indicators (Kase Index, Brent oil, CDS, and US real rate) are subsequently verified for compliance with these conditions. Therefore, this ensures a link between high-frequency financial data and low-frequency fundamental macroeconomic variables.

In the next step, the behavior of assets within portfolios across different macro regimes is explored. Using the daily returns of each component, the averages and volatilities in each model are calculated. Next, the portfolio is optimized separately for each model, taking into account those factors:

1. Minimizing the ES (ES, 97.5%).
2. Maximizing expected returns.
3. Regulatory restrictions of the NBK (share of gold + liquidity + investment $\geq 64\%$, as well as individual restrictions for shares of strategic, EM, and AI portfolios.

The formulation of the optimization task reflects the priorities of the reserves management: the safety of capital is more necessary than yield, which is reflected in the high “risk price”, the portfolio must be resilient to extreme market movements, and the structure ought to remain within imposed regulatory boundaries.

The standard optimization uses historical assessment of volatility, which indicates there is a constant risk over the given period. However, emerging markets like Kazakhstan are characterized by volatility clustering, with periods of calm alternating with turbulence periods.

In order to account for this feature, this study applies the GARCH model, which generates conditional forecasts of volatility and covariance dynamics. This approach ensures forward-looking optimal weights that take into account those features with higher expected risk in slowdown regimes, lower risk in expansion regimes, and stronger sensitivity to recent market shocks. The GARCH approach confirms the conclusions of the historical model and enhances the “flight to safety” effect by increasing the gold’s share during economic slowdown.

Finally, the constructed model helps to form the portfolio, which is adaptive to market conditions, lowers the loss probability, accounts for the macroeconomic structure of Kazakhstan (oil dependence and sensitivity to external rates), and justifies changes in the portfolio structure not solely by relying on current market conditions, but also from the fundamental dynamics.

5. Results

Portfolio replication results and fundamental risk characteristics

According to the methodology, the first step in the analysis is to construct a replication of the structure of Kazakhstan's international reserves using market proxy indices for the six components of the GXFA portfolio: gold, liquidity, strategic, investment, AI, and EM. The replication is computed in the proportions that correspond to the actual data of the NBK, which made it possible to replicate the dynamics of the total market risk of the portfolio and assess its behavior under different market conditions.

The distribution of 10-day portfolio returns revealed pronounced asymmetry in the presence of heavy tails:

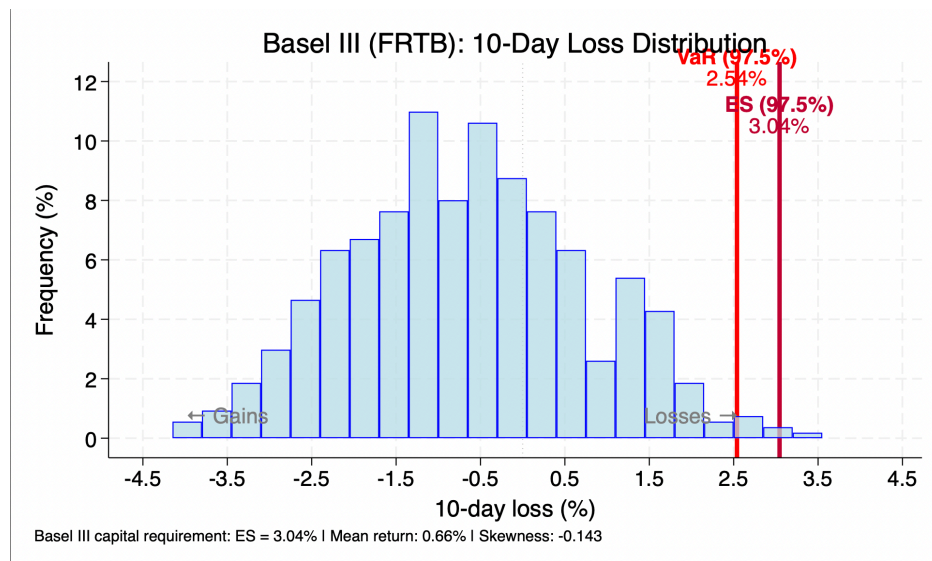


Figure 2. Basel III: 10-Day Loss Distribution (Current Portfolio)

The histogram reflects the empirical distribution of 10-day portfolio value changes, calculated using historical modeling. This is the actual data that characterizes how the portfolio behaved over the given period. The two vertical lines are plotted on top of the empirical distribution, reflecting GARCH estimates of VaR and ES that follow the requirements of Basel.

Most of the data lies in the range from -1.5% - +1%. This is typical for reserve portfolios, where a significant part of the assets are highly liquid and relatively low-risk instruments, which smooths out extreme changes in prices. The historical histogram smooths out spikes in risk over the period due to the fact that each 10-day period has an equal value. GARCH estimates, on the contrary, reinforce the impact of recent periods of high volatility, take into account volatility clustering, and

adapt to market changes in real time. Hence, the GARCH lines are above and to the right of the levels that could be taken if we used a solely historical method. This highly corresponds with the Basel rationale, regulators prefer forward-looking, not backward-looking risk assessment. The rise of ES relative to VaR of about 20% corresponds to the features of the heavy-tailed distribution.

Daily VaR/ES dynamics

For assessing the market risk of NBK's GFXA, daily indicators of VaR and ES were calculated at the confidence levels of 95.5% and 97.5% using two approaches – Historical Simulation and GARCH (1,1). Since both approaches provide different perspectives on the return distribution, comparing them helps to understand how risk evolves depending on volatility changes, tail behavior, and the market regimes.

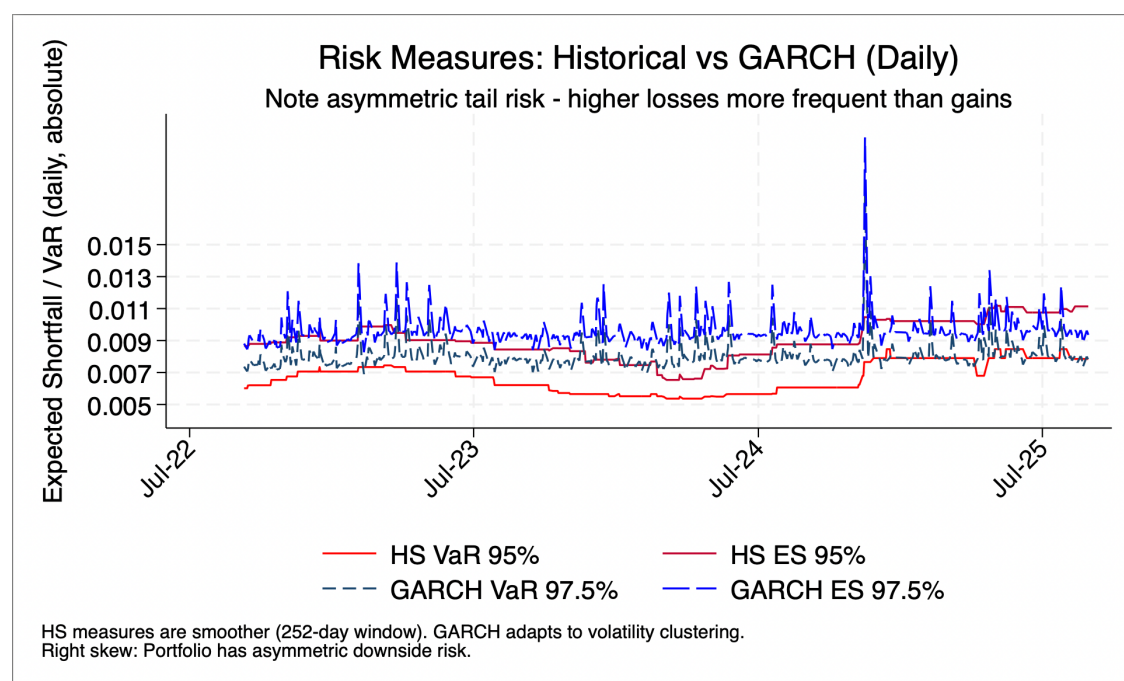


Figure 3. Risk measures: HS vs GARCH (Daily)

The graph above shows the HS-VAR (95%) fluctuates in a relatively narrow range – 0.55% to 0.80%, whilst GARCH-VaR could reach 0.90-1.20%, in real cases – hitting 1.80%, which is visually any HS values. A similar situation is observed for ES: HS-ES (95%) has generally losses from 0.80 to 0.95%, whereas GARCH-ES (97.5%) is stable at a little more than 1.0%. Sometimes, in high volatility periods, it could reach 1.70-2.10%

Summary statistics						
Measure	Mean	Std Dev	Min	P50	P95	Max
VaR95_HS_r_total	0,669	0,094	0,537	0,670	0,790	0,846
ES95_HS_r_total	0,903	0,118	0,654	0,888	1,110	1,117
VaR GARCH (97.5%) - 1-day	0,803	0,082	0,690	0,784	0,959	1,815
ES GARCH (97.5%) - 1-day	0,963	0,098	0,827	0,941	1,149	2,171
ES GARCH (97.5%) - Basel 10-day	3,044	0,311	2,614	2,976	3,633	6,866

Table 2. Summary statistics of VaR and ES

Conclusion: GARCH models show that risk levels exceed historical estimates by around 25-35% in calm periods and even by 100-150% during stress periods. Figure 3 shows sharp spikes in ES, which are practically not observed in HS methods. For example, in the market, when gold experienced a sharp correction, GARCH ES (97.5%) rose from 1.0% to 1.5%. In addition, in the beg-24, beginning of 24obal volatility intensified, the figures reached 2.0-2.2%, whilst the HS model did not react much to these events, so it barely reached 1.1%.

The above-mentioned results indicate the following two facts: the portfolio's return distribution displays significantly heavier tails than those assumed under a normal distribution, and HS is inherently slow-moving and tends to smooth volatility, whilst the GARCH models are able to detect short-term risk bursts.

Backtesting

Backtesting of VaR and ES models was conducted to check the quality of risk assessment and compliance with the standards of Basel. The analysis of the exceedance rate of VaR showed that with the confidence level of 97.5%, the actual frequency of violations was about 3.3%, which was a bit higher than the theoretical expected level of 2.5%. This means that the GARCH model sometimes underestimates the risk of the individual tail events, in particular during the periods of gold price spikes. In addition, the indicator of estimation risk was computed – the degree of variability of ES scores relative to their average level. For the GARCH model, the coefficient of variability was around 10%, which reflects moderate sensitivity of predicted ES to short-term volatility spikes. This corresponds to the fact that the model reacts quickly to local bursts of uncertainty but does not illustrate excessive instability. Moreover, model specification risk was

assessed, defined as the divergence between the historical ES estimate and model-generated ES. The average ratio of HS-ES to GARCH-ES is about 1.10, which means that HS gives about 10% higher risk estimates in calm periods, while GARCH becomes the dominant source of risk signals in times of economic slowdowns. Putting everything together, the backtesting results show that GARCH provides a realistic and fairly conservative assessment of market risk, but requires adjustments for tail losses. It is for this reason that the tail-loss correction factor is further applied, which makes it possible to transform GARCH-ES into a more stable and regulatory-compliant assessment.

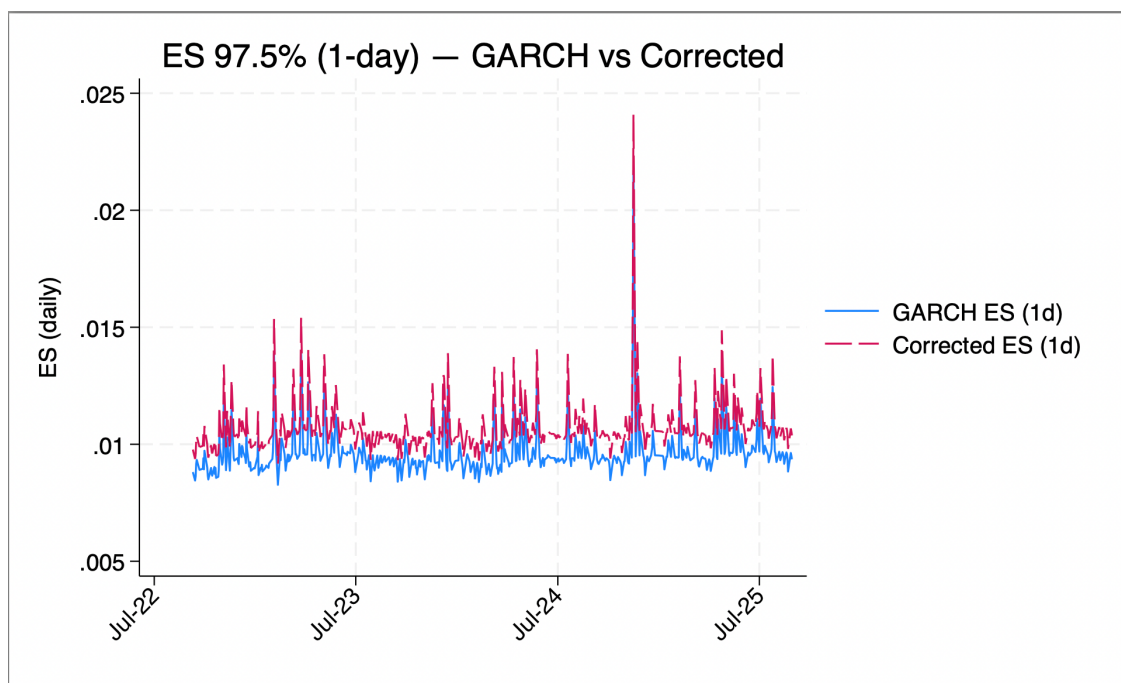


Figure 4. ES 97.5% (1-day) – GARCH vs Corrected

Figure 4 compares the basic GARCH ES indicator and the correct option, which was enhanced using the empirically observed tail-loss coefficient. This approach makes it possible to assess how much the model ES corresponds to the realized losses associated with VaR exceedances during the sample period.

For the final data point in the sample, the GARCH model estimated the one-day ES (0.936%), while empirically observed average loss associated with the same exceedance events was 1.038%. The difference in absolute terms might look minor – only 0.10 percentage points, however in relative terms it corresponds to a 10.9% underestimation – a substantial difference in terms of

capital adequacy assessment. Put differently, it means each realized loss was deeper by almost 11% compared to the predicted model.

For financial instruments with clear asymmetry in the distribution of returns, especially portfolios containing gold, such underestimation is in fact typical: The GARCH model carefully takes into account volatility clustering.

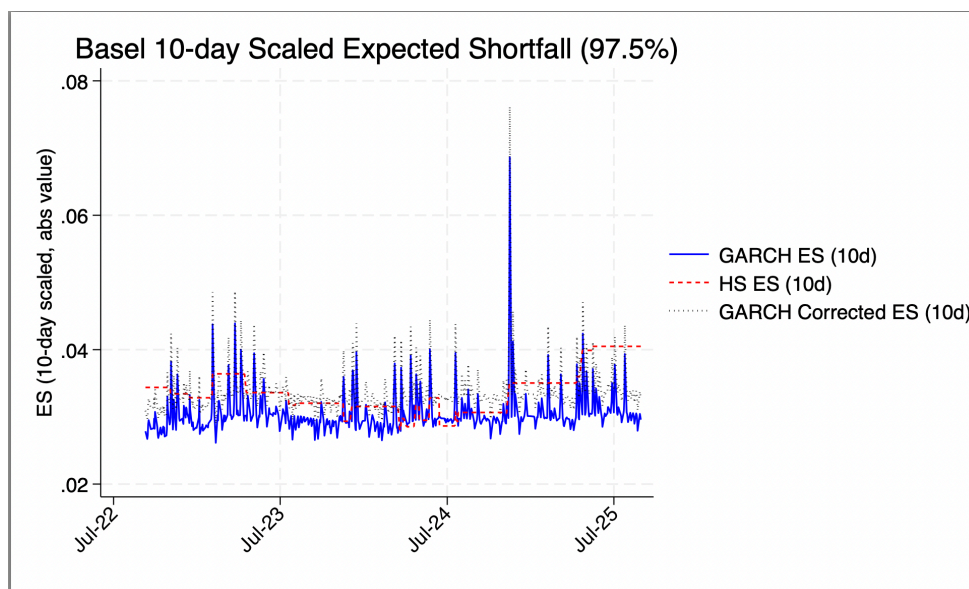


Figure 5. Basel 10-day scaled ES (97.5%)

The Basel standard requires the ES calculation on 10 trading days, which reflects potential losses in case of forced liquidation of positions. According to the last sample data, the following values are obtained: GARCH-ES (10-day) = 2.96%, HS-ES (10-day) = 3.4%, corrected GARCH-ES (10-day) = 3.4-3.5%. It means that during calm macroconditions, HS yields higher risk estimates than GARCH. However, this relationship reverses drastically during periods of market stress. For instance, on July 24, GARCH-ES (10-day) reached 6-7%, while HS-ES (10-day) had around 3.5-4%. In fact, these all mean that during economic slowdowns, 10-day ES is doubled, which reflects: the spillover of volatility clusters into medium-term dynamics, rising correlations across the portfolio's components, and greater sensitivity of the portfolio to gold, which in stress periods can display both upward jumps and sharp corrections. Therefore, the 10-day GARCH-ES is more realistic when it comes to the assessment of potential losses during the shock period.

Markov regime switching

This section presents the application of the Markov Regime Switching model (MRS) to identify the phases of Kazakhstan's economic cycle and assess how different asset classes of the portfolio respond to changes in the variance structure of returns. The MRS approach allows us to statistically divide the sample into periods of "Economic expansion" and "Economic slowdown", where each regime is characterized by its own dynamics of volatility and asset behavior. The model was evaluated based on data from daily portfolio returns for the period 09/09/2022–08/29/2025, which includes 546 observations.

The model was evaluated according to the following scheme: (1) Optimization algorithms - the model automatically switched between optimization methods, such as BHHH — used when the log-likelihood surface exhibited strong curvature and BFGS — used for the final, more precise approximation of the maximum, (2) Sample with gaps - each data gap triggered a restart of the conditional distribution, preventing the accumulation of errors in the conditional volatility, and (3) Convergence criterion - Iterations continued until the log-likelihood value stabilized at -650.27844 .

The division of Kazakhstan's economy into two regimes is identified as follows:

Regime	Economic Interpretation	Share of days	# of observations
Expansion	Expansion, low and steady volatility	82%	448
Slowdown	Slowdown increased volatility, increased risks	18%	98

Table 3. The division of Kazakhstan's economy

Thus, the vast majority of the period is characterized by a phase of stable growth, but about a fifth of the sample represents periods of stress, where a "slowdown mode" is manifested.

For each asset, its own GARCH(1,1) model was evaluated within a specific. This makes it possible to detect a change in the dispersion structure, which is the main nature of mode switching.

The table of conclusions is abbreviated below.

Portfolio type	μ (Expansion)	σ (Expansion)	μ (Slowdown)	σ (Slowdown)	Description
Gold	0.11%	2.65%	0.07%	0.81%	Safe-haven stabilization: Volatility drops 70% in slowdown → gold

					becomes predictable defensive anchor
Liquidity	0.008%	0.05%	0.01%	—	Volatility is minimal in both regimes
Strategic	0.02%	0.71%	0.04%	0.26%	Volatility decreases 63% in slowdown as credit markets freeze with wider but stable spreads
Investment	0.017%	0.47%	0.018%	0.62%	Moderate response; volatility increases in slowdown
AI	0.048%	2.07%	0.07%	3.70%	Alternatives sharply amplify volatility
EM	0.021%	0.68%	−0.01%	1.08%	EM deteriorates in slowdown — both in risk and returns

Table 4. Interpretation of Regime-Specific Results

Regime-Specific asset behavior analysis

The analysis of the regime indicates that there are a number of trends that dictate the best structure of the reserve portfolio of Kazakhstan. First, gold affirms it to be a safe haven. Its volatility is greatly decreased to of 2.65% in growth mode and to 0.81% in stress mode in the average yield, even in the slowdown phase. This implies that at times when market conditions are at their worst, gold is rendered. A more predictable and constant asset, and it should have a higher weight in the portfolio in the case of an economy. Goes to a slowdown mode.

Strategic bonds show higher volatility during Expansion (0.71%) than Slowdown (0.26%) — a 63% decrease. This counterintuitive pattern occurs because during quiet periods, credit spreads are compressed and sensitive to sentiment shocks. In slowdown, credit markets 'freeze' with wider spreads but more stable pricing.

Third, emerging markets are behaving poorly in a slowdown mode: average returns are turning negative, and volatility is almost doubling. This makes EM assets undesirable during a stressful period.

Finally, alternative investments significantly increase the overall volatility of the portfolio. In slowdown mode, there is a sharp jump in volatility to 3.7%, which confirms that alternative assets are not protective, and their share in the reserve portfolio should be minimal.

Covariance matrix construction approaches

Three approaches were used in the study to construct covariance matrices. The first method is based on a combination of GARCH volatility and historical correlations. In this case, the covariance matrix is formed as:

$$\Sigma = D_{GARCH} R_{hist} D_{GARCH},$$

This ensures the stability of the estimate and avoids the use of multidimensional GARCH models, known for their complexity and poor convergence. The second approach, DCC-GARCH, involves modeling dynamic correlation. However, attempts to evaluate the DCC-GARCH model(1,1) for six assets encountered serious difficulties: the algorithms did not converge in conditions of a short sample, high mutual correlation of assets, and the presence of gaps in the data. In this regard, this method was abandoned and a more stable first approach was adopted.

Mode-dependent GARCH is the third method, which entails the development of individual covariance matrices for each mode: Σ_1 For Expansion and Σ_2 For Slowdown. These matrices reflect that they vary in the amount of volatility as well as the structure of correlations, and are applied further in portfolio optimization.

Historical simulation optimization results

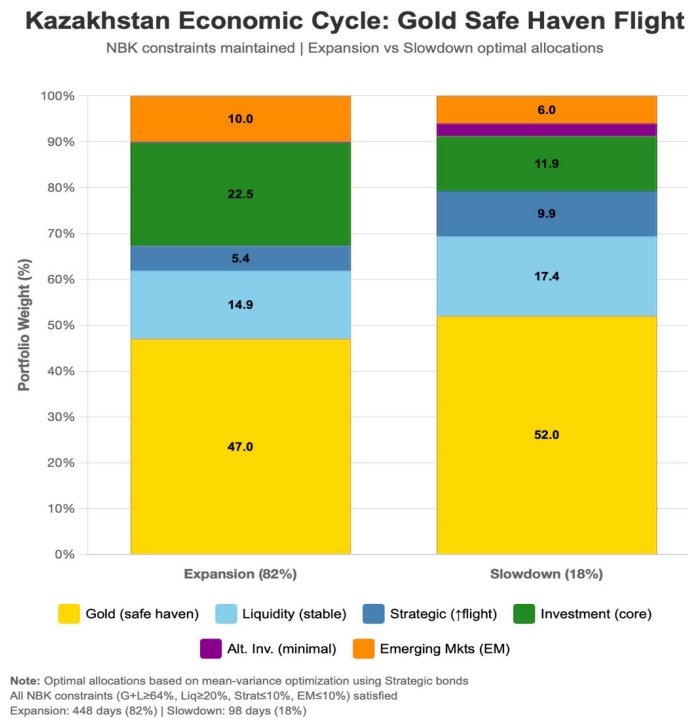


Figure 6. Optimal portfolio allocations across economic regimes: Expansion vs. Slowdown (HS)

The Expansion regime with low volatility and stable returns occupies 82% of the observations. Under the Historical Simulation approach, which uses unconditional mean-variance estimates based on 252-day rolling windows, the optimization assigns gold a weight of approximately 47.0%, liquid assets – 14.9%, strategic bonds – 5.4%, investment securities – 22.5%, alternatives - 0.2%, and EM markets - 10.0%. These findings indicate that during calm periods, the historical model permits a more balanced allocation across asset classes, though with notably higher exposure to investment-grade bonds and emerging markets.

The structure changes in the Slowdown regime that occupies approximately 18% of the sample. Under Historical Simulation, gold increases to 52.0%, liquidity to 17.4%, strategic assets increase to 9.9%, investment decreases to 11.9%, alternatives to 2.8%, and EM to 6.0%. The weight of gold has increased by 5.0 percentage points, and the EM percentage has fallen by 4.0 percentage points due to rising risks and heightened uncertainty during the slowing regime.

GARCH optimization results

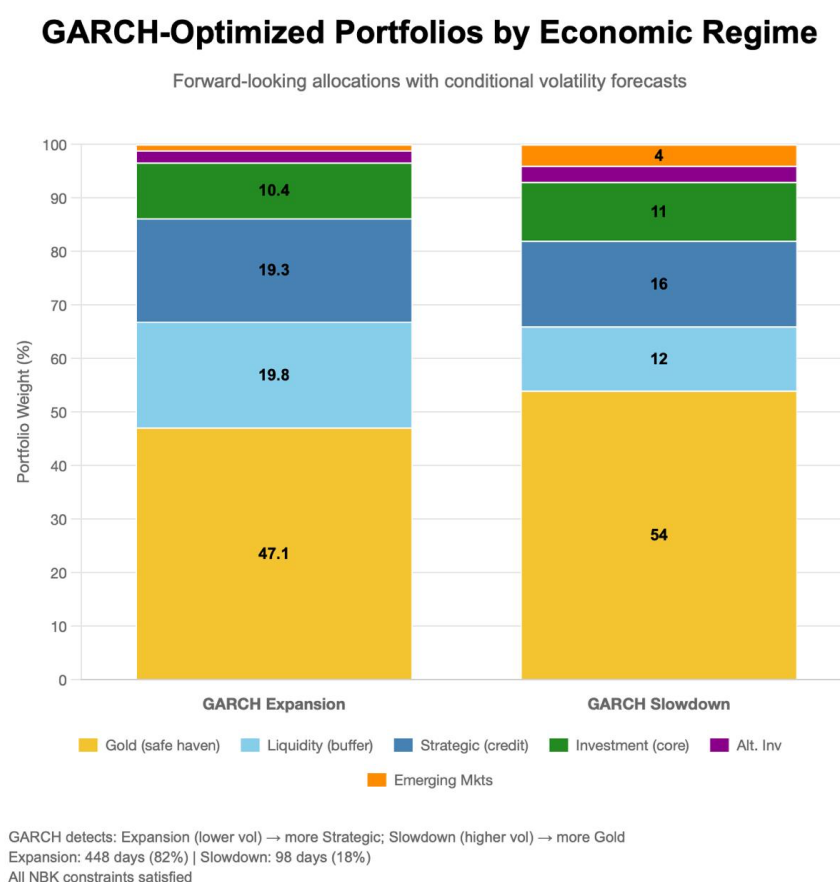


Figure 7. GARCH-Optimized Portfolios by Economic Regime

The findings show that under the two mentioned regimes, namely, economic expansion and slowdown, the behavior of assets differs significantly when conditional volatility is properly modeled using GARCH(1,1), marking a shift in the optimal weights compared to Historical Simulation.

Expansion regime (GARCH)

In the expansion mode, when market volatility remains relatively low and returns on most assets are stable, the GARCH model allows for a more sophisticated risk allocation compared to Historical Simulation. Under such conditions, the optimal weight of gold is about 47.1%, nearly identical to the Historical Simulation result, but with significantly different allocations across other asset classes.

The GARCH model allocates 19.8% to liquidity (versus 14.9% in HS), 19.3% to strategic bonds (versus 5.4% in HS), 10.4% to investment securities (versus 22.5% in HS), 2.3% to alternatives (versus 0.2% in HS), and 1.1% to EM (versus 10.0% in HS). This reflects GARCH's sensitivity to the conditional volatility structure, which recognizes that in calm periods, strategic credit instruments offer superior risk-adjusted returns when volatility clustering is properly accounted for.

In quiet periods, alternative asset classes — strategic bonds, investment securities, and even parts of the liquidity sector — offer more favorable risk-return ratios when their time-varying volatility is correctly modeled. Gold in such periods shows moderate volatility, but at the same time generates less income compared to the bond component. In this mode, gold plays an auxiliary role — it reduces the overall correlation of the portfolio, but is not the main driving force behind its effectiveness.

Slowdown regime (GARCH)

A markedly different picture is observed in the slowdown mode. In the phase of increased uncertainty, credit and market risks are sharply increasing, volatility indicators of strategic and emerging markets are rising, and alternative assets are showing sharp and dispersive movements. Against this background, gold becomes the most stable asset in the entire GFXA structure: its volatility decreases dramatically from 2.65% to 0.81%, while the yield remains moderately positive.

This countercyclical behavior causes the GARCH optimization model to recommend a larger percentage of gold at 54.0% (versus 52.0% in Historical Simulation), indicative of a heightening

of its risk mitigation effect and a more prominent role in mitigating tail risk. Liquidity is allocated at 12.0% (versus 17.4% in HS), strategic bonds at 16.0% (versus 9.9% in HS), investment at 11.0% (versus 11.9% in HS), alternatives at 3.0% (versus 2.8% in HS), and EM at 4.0% (versus 6.0% in HS).

Thus, gold becomes an important instrument that enables NBK to stabilize the portfolio when conditions become turbulent, serving as a market stability anchor.

It is necessary to point out that the rise in the optimum weight of gold does not happen as a result of an increase in its profitability, but rather due to a relative decline in the risk profile of other underlying asset classes. Emerging markets begin to show negative average returns with a sharp increase in volatility, the strategic component loses its attractiveness due to increased credit stress, and alternative investments become a source of disproportionate surges. All this makes gold the most reliable source of stabilization and reduces the appeal of risky investments.

Comparing HS and GARCH approaches

Asset Class	HS Expansion	GARCH Expansion	Change (pp)	HS Slowdown	GARCH Slowdown	Change (pp)
Gold	47.0%	47.1%	+0.1	52.0%	54.0%	+2.0
Liquidity	14.9%	19.8%	+4.9	17.4%	12.0%	-5.4
Strategic	5.4%	19.3%	+13.9	9.9%	16.0%	+6.1
Investment	22.5%	10.4%	-12.1	11.9%	11.0%	-0.9
Alternative Inv.	0.2%	2.3%	+2.1	2.8%	3.0%	+0.2
Emerging Markets	10.0%	1.1%	-8.9	6.0%	4.0%	-2.0

Table 5. Comparison of HS vs. GARCH optimization results

Additional analysis showed that optimization based on the GARCH volatility estimate reinforces the differences between the regimes: while Historical Simulation smooths out the differences between periods of growth and slowdown due to its unconditional, backward-looking nature, GARCH, on the contrary, captures short-term spikes in risk and, accordingly, increases the weight of gold more aggressively during stressful periods (54.0% versus 52.0% in HS).

The most important result is that the GARCH approach significantly enhances the differences between the regimes. If Historical Simulation smooths out the risks by averaging over fixed windows, then GARCH reveals the real picture: during a slowdown, the actual expected volatility turns out to be much higher than the unconditional average. This naturally leads to a more

pronounced increase in the weight of gold (+2.0pp from expansion to slowdown in GARCH versus +5.0pp in HS), a more dramatic reduction in the share of EM (4.0% in GARCH slowdown versus 6.0% in HS slowdown), and a more sophisticated treatment of strategic bonds based on their regime-dependent volatility behavior (16.0% in GARCH slowdown versus 9.9% in HS slowdown).

This highlights the importance of dynamic models for portfolios with a high concentration of tail risk.

Key insights from MRS analysis

The analysis indicates that the economy of Kazakhstan is characterized by a strong two mode structure: calm periods (82%) are observed to alternate with sharp periods of high volatility (18%). This dynamic is characteristic of resource-based economies that tend to be shock sensitive to external impacts like commodity prices, sanctions restrictions, or massive capital flows.

Gold is an important stabilizer of the portfolio. It demonstrates a negative or low correlation with internal financial shocks, significantly reduces volatility in stress mode, and maintains a positive average return. That is why both the Historical Simulation and GARCH optimization models increase the weight of gold in the slowdown mode, though the GARCH model responds more dynamically to recent volatility conditions.

Strategic bonds are losing their appeal as they slow down. Their volatility decreases, but at the same time, their profitability decreases, and correlations with other risky assets increase. This leads to a decrease in their share in the portfolio during the transition to a slowdown.

Emerging markets and alternative assets turn out to be the riskiest: they show increased volatility, lower or negative returns, and increased tail risks. Therefore, GARCH optimization almost always tends to dramatically minimize its weight.

The Markov regime switching model demonstrates that Kazakhstan's economy has a stable dual-mode structure: a dominant growth phase and a critically important slowdown phase, in which the main portfolio risks are formed. The dynamics of assets in the GARCH model differ significantly between the modes, which confirms the inefficiency of using a constant covariance matrix.

Gold turns out to be a central stabilizing element, reducing volatility and improving ES performance during a slowdown. The risks of emerging markets and alternative instruments increase significantly with the transition to the second mode. GARCH optimization reveals a

pronounced "flight into quality" in the slowdown phase and an expansion in the share of credit strategies during the growth period.

Thus, the use of MRS-GARCH structures significantly increases the realism of risk management and allows us to create portfolios adapted to the empirical volatility environment and recent market dynamics, rather than to average historical parameters. The combination of both approaches provides complementary perspectives: Historical Simulation offers stable, long-term benchmarks, while GARCH provides forward-looking, volatility-adaptive tactical guidance.

Optimization outcomes: regime-dependent optimal gold allocation

Optimization of the reserve portfolio structure based on the Markov Regime Switching dual-mode model makes it possible to quantify how the optimal share of gold changes depending on the state of the financial market and the fundamental level of risk. The primary objective of the optimization the purpose of which is to calculate a reasonable amount of gold, which will reduce the market risk, quantified by ES, without being so vulnerable to shocks because of its portfolio.

The results demonstrate that the optimal share of gold is not fixed but rather dynamic and regime-dependent. In the slowdown regime, the optimal percentage of gold rises to 54.0% under GARCH optimization (52.0% under Historical Simulation), constituting the primary risk mitigation mechanism. The optimal share is lower during expansion mode at approximately 47%, which reflects the reduced demand for protective assets and the increased significance of return-generating instruments during stable periods. This confirms the relevance of a flexible methodology in establishing the structure of reserves, and the appropriateness of bringing in tactical modifications to the allocation of gold.

Taken together, the results show that the prominent role of gold in the structure of Kazakhstan's reserves is not accidental: optimization based on strict quantitative methods confirms its critical role in reducing expected extreme losses. However, the analysis also reveals an important nuance: over long periods of stability, excessive concentration in gold creates opportunity costs in the form of foregone returns. This means that in quiet periods, it is possible to gradually reduce the share of gold in favor of strategic and investment assets with higher returns, but at the same time with the obligatory preservation of a sufficient protective core that will perform a countercyclical function in the event of a sharp deterioration in market conditions.

The outcomes of the results indicated by the model therefore propose a policy that is balanced and regime-sensitive in the management of the gold allocation: a moderate level around 47% during

expansion and a stable rise to 54% during slowdown mode. This establishes the foundation of an adaptive reserve management strategy that integrates the need for reliability, liquidity, and market shock resistance, while minimizing long-term opportunity costs resulting from being overly concentrated in a single asset.

6. Conclusion

This study provides a thorough evaluation of the market risk of Kazakhstan's GXFA using contemporary stress testing tools and parametric volatility models. The obtained results can be effectively used to create a comprehensive perspective on risk dynamics, the resilience of the portfolio to shocks, and opportunities to improve the management of reserve asset structures.

First, the replication of the GXFA portfolio demonstrated that the returns distribution is highly asymmetric and skewed, with pronounced heavy tails. This points to the weakness of classical historical methods with regard to extreme losses. The GARCH(1,1) model is able to capture the clustering effect of volatility, which enables short-term risk spikes to be identified—effects that are smoothed out by the historical simulation process due to its backward-looking, unconditional nature. A comparison of daily VaR and ES values showed that GARCH estimates risk 25-35% higher than Historical Simulation during calm market periods and 100-150% higher during turbulent periods.

Second, the backtesting results indicate that at a confidence level of 97.5%, the actual frequency of VaR violations was approximately 3.3%, slightly higher than the theoretical expected value of 2.5%. This suggests that the probability of extreme events, particularly those involving gold price movements, is underestimated by standard models. Given that the gap between realized and projected ES values amounted to 10-12%, the necessity for tail risk adjustments is confirmed. Further, estimation risk and specification risk indicators were evaluated: the former reflects the high sensitivity of ES to short-term volatility spikes, while the latter reveals a persistent bias between historical and model-based estimates in stable regimes.

Third, the Markov Regime Switching analysis revealed that the composition of optimal portfolios changes substantially depending on the stage of the economic cycle. The relative attractiveness of gold as a protective asset increases significantly during slowdown periods (optimal allocation: 54.0% under GARCH, 52.0% under Historical Simulation), compared to reduced attractiveness of strategic assets and highly risky segments. In contrast, maintaining a high percentage of gold during the expansion phase generates opportunity costs in the form of foregone returns, as cyclical assets offer superior profitability during stable periods. The regime-dependent approach makes it possible to shape a more adaptive strategy for managing reserves that accounts for changing market conditions.

Fourth, the optimization analysis revealed that the optimal proportion of gold in the portfolio should not be fixed but rather dynamic and regime-dependent. During slowdown regimes, the optimal percentage of gold rises substantially and constitutes the primary risk mitigation mechanism. The optimal share is lower during expansion mode (approximately 47%), which reflects reduced demand for protective assets and increased significance of return-generating instruments. Importantly, the GARCH-based optimization produces more pronounced regime differences than Historical Simulation: GARCH increases gold to 54.0% in slowdown (versus 52.0% in HS) and dramatically reallocates strategic bonds (19.3% in expansion, 16.0% in slowdown under GARCH, versus 5.4% and 9.9% under HS). This confirms the relevance of forward-looking, volatility-adaptive methodologies in establishing the structure of reserves and the appropriateness of tactical modifications to gold allocation.

Overall, the outcomes of this research demonstrate that more sophisticated risk assessment tools that account for the nonlinear nature of return distributions and regime-dependent volatility transitions are essential for effective reserve management. GARCH models with tail risk adjustments provide more realistic estimates of VaR and ES compared to unconditional historical methods. The combination of Markov Regime Switching models and dynamic optimization processes enhances the effectiveness of structural decisions in the management of international reserves.

The results may be utilized to refine the approach to market risk measurement and to inform strategic and tactical asset allocation decisions at the National Bank of Kazakhstan. Specifically, the findings support a regime-sensitive gold allocation strategy: maintaining approximately 47% during stable expansion periods to minimize opportunity costs, while increasing to 54% during slowdown periods to maximize tail risk protection. This adaptive framework integrates the need for reliability, liquidity, and market shock resistance while balancing long-term return objectives.

7. References

1. Acerbi, C., & Székely, G. (2014). Backtesting expected shortfall. *Risk Magazine*, 27(11), 76–81.
<https://www.academia.edu/>
2. Bali, T. G., Mo, H., & Tang, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *Journal of Banking & Finance*, 32(2), 269–282.
<https://www.sciencedirect.com/science/article/abs/pii/S0378426607001823>
3. Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? *Financial Review*, 45(2), 217–229.
<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6288.2010.00244.x>
4. Beckmann, J., Berger, T., & Czudaj, R. (2018). Gold price dynamics and the role of uncertainty. *Quantitative Finance*, 19(4), 663–681.
<https://www.tandfonline.com/doi/full/10.1080/14697688.2018.1508879>
5. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
<https://www.sciencedirect.com/science/article/abs/pii/0304407686900631>
6. Chen, C. Y.-H., & Chiang, T. C. (2015). Empirical analysis of the intertemporal relationship between downside risk and expected returns: Evidence from time-varying transition probability models. *European Financial Management*, 21(4), 673–705.
<https://onlinelibrary.wiley.com/doi/full/10.1111/eufm.12079>
7. International Monetary Fund. (2003). Guidelines for foreign exchange reserve management. IMF.
<https://www.imf.org/external/np/mae/ferm/2003/eng/part2a.pdf>
8. International Monetary Fund. (2020). Guidelines for foreign exchange reserve management. IMF.
<https://www.elibrary.imf.org/display/book/9781589062603/ch004.xml>

9. Lazar, E., & Zhang, N. (2019). Model risk of expected shortfall.
Journal of Banking & Finance, 105, 74–93.
<https://www.sciencedirect.com/science/article/pii/S0378426619301220>
10. McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for
heteroscedastic financial time series.
Journal of Empirical Finance, 7(3–4), 271–300.
<https://www.sciencedirect.com/science/article/abs/pii/S0927539800000128>